

BIVARIATE (CONT.) AND REVIEW

Zack Kertcher

Statistics for Management
Fall 2016

Plan for today

1. Class logistics
2. Presentations
3. Bivariate relationships (cont.)
4. Mid-term review

PRESENTATIONS

BIVARIATE RELATIONS (CONT.)

What we already know and what need to know

Relationship	Statistics	Plots
F / F	<code>table</code> <code>prop.table(t)</code> <code>margin.table(t)</code>	<code>mosaicplot</code> <code>barplot()</code>
N / F		<code>boxplot()</code>
N / N		<code>plot(x~y)</code> <code>abline(lm(x~y))</code>

What we already know and what need to know

Relationship	Statistics	Plots
F / F	<code>table</code> <code>prop.table(t)</code> <code>margin.table(t)</code>	<code>mosaicplot</code> <code>barplot()</code>
N / F	<code>aggregate</code> <code>describeBy</code> <code>group_by()</code>	<code>boxplot()</code>
N / N	<code>cor</code> <code>lm()</code>	<code>plot(x~y)</code> <code>abline(lm(x~y))</code> <code>pairs()</code> <code>Corrplot()</code>

EXAM REVIEW

Part I

1. From Blackboard, reload the real estate data as **realestate**.
2. Prepare variables as needed. (Hint: make sure that variables are in the type they need to be, and that all variables are ready for analysis)
3. What is the 75th percentile of price? What are the 10th and 90th percentiles of lotsize?
4. Examine the following variables *individually*: lotsize, price, stories. In your examination, use univariate statistics, and plots.
5. Discuss each variable, including the distribution shape, central tendency, skew, and other observed features.

Part II

1. How much, on average, does a house with air conditioning costs, compared to one without it (airco)?
2. What is the difference in the distribution of price between houses with and without air conditioning? Use relevant statistics and plots.
3. Using the correlation function, find what is the relationship between price and lotsize.
4. Use a correlation matrix to examine the relationships across all numeric variables in the data. Which variables are correlated? (Hint: assume that values over 0.4 or under -0.4 indicate a correlation). What are the two highest correlations in the data?